

ADM 2304 -- ASSIGNMENT 1

Due Date: Tuesday, February 9, 2016 at midnight.

Instructions:

1. You may use Minitab or other software for any calculations. However, you must show your manual calculations when asked. You may paste your output onto your assignment to show your use of Minitab; however, this output does not replace any of the steps outlined below. This means that answers that are exclusively Minitab output may receive only one mark.
2. If you are performing a hypothesis test, make sure you state the hypotheses, the level of significance, the decision rule in terms of the critical value, the test statistic or p-value, your decision (whether to reject or not to reject the null hypothesis), and your conclusion in managerial terms. These steps must be completed in addition to any Minitab output.
3. The data in the Minitab data files can be found in the file **Assign1Data.xls**.
4. Remember to include your integrity statement. Your assignment should be uploaded to Blackboard in pdf format.

1. [10 marks]

In the 2015 federal election, 39.47% of the electorate voted for the Liberal party, 31.89% for the Conservative party, 19.71% for the NDP, 4.66% for the Bloc Quebecois, and 3.45% for the Green party.

- (a) A recent poll of 1500 respondents found that 28% support the Conservative party. Test whether this is sufficient evidence to conclude that the level of support has dropped since the election. Use the 5% level of significance and show your manual calculations.

Test and CI for One Proportion

Test of $p = 0.3189$ vs $p < 0.3189$

Sample	X	N	Sample p	95% Upper Bound	Z-Value	P-Value
1	420	1500	0.280000	0.299069	-3.23	0.001

4 marks, 1 for each point below:

-Ho: $p=0.3189$; Ha: $p < 0.3189$

-z = $(0.28 - 0.3189)/\sqrt{0.3189*0.6811/1500}=-3.23$

-rejection region is $Z < -1.645$ or p-value = $P(Z < -3.23) = 0.001$

-decide to reject Ho and conclude support has dropped.

- (b) Suppose you want to estimate the national level of support for the Conservatives using a 95% 2-sided confidence interval with a margin of error of $\pm 1\%$. What sample size would be required?

$M = 0.01$, $z = 1.96$ for 95% CI, $p\text{-hat} = .28$, $q\text{-hat} = .72$

$$n = pq (z / M)^2 = 0.28*0.72 * (1.96/0.01)^2 = 7745$$

2 marks:

-1 mark for z-value and M value

-1 mark for use of proper formula and calculation

If the solution uses the conservative values of $p=q=0.5$ or $p=.3189$, then only 1 mark for the z-value and M, but no second mark.

- (c) Suppose that, in a random sample of 17 University of Ottawa students, only 2 indicated a preference for the Conservatives. Test whether this is sufficient evidence to indicate that the level of support for the Conservatives among U of O students is lower than the 31.89% share of the popular vote in 2015. Use the .05 level of significance and show what probability you need to find to calculate the p-value for the result.

Test and CI for One Proportion

Test of $p = 0.3189$ vs $p < 0.3189$

Sample	X	N	Sample p	95% Upper Bound	Z-Value	P-Value
1	2	17	0.117647	0.246180	-1.78	0.037

*** NOTE * The normal approximation may be inaccurate for small samples.**

Minitab will calculate the proper p-value, using the binomial probability:

Test and CI for One Proportion

Test of $p = 0.3189$ vs $p < 0.3189$

Sample	X	N	Sample p	95% Upper Bound	Exact P-Value
1	2	17	0.117647	0.326193	0.057

4 marks:

1 for hypotheses of $H_0: p=.3189$, $H_a: p<.3189$

1 for recognizing normal approximation **not appropriate** since we only observe 2 supporting Conservatives or $np = 17*.3189=5.42 < 10$.

1 for calculating p-value of 0.057 (whether using Minitab or binomial calculation $P(X \leq 2) = 0.057$, based on $n=17$ and $p=0.3189$)

1 for decision not to reject H_0 and conclusion that there is insufficient evidence to conclude support lower than .3189.

Note that if the solution uses the z-statistic and the normal approximation to find the p-value, then give 2 marks only, for hypotheses and decision to reject/conclusion there is sufficient evidence.

Some students may make the adjustment from section 11.6, using $p\text{-hat} = (2+ 2)/(17+4)=0.19$.

The z-statistic would be $(0.19 - .3189)/\text{sqrt}(.3189*.6811/21) = -.1289/.1017 = -1.267$, with a p-value of $P(z < -1.267) = 0.102$.

A 95% CI would be an upper bound of $0.19 + 1.645 * \text{sqrt}(.3189*.6811 / 21)$ or $0.19 + 1.645*0.1017$ or $0.19 + 0.1673 = 0.357$, which is closer to the binomial calculation above.

This solution can get one more mark for a total of 3 out of 4, if the confidence interval estimate is given without a p-value.

Give full marks of 4 out of 4, if the p-value of 0.102 is given instead of 0.057.

2. [10 marks]

The file **bmisamples.mtw** contain the BMI of two samples of men and women. Two of the columns, **OWmale** and **OWfemale** code the BMI values as:

0 - if BMI \leq 25.4 (these are considered “not overweight”);

1 - if BMI \geq 25.5 (these are considered “overweight”).

- Test whether there is sufficient evidence to show that the proportion of overweight males (proportion of males who are overweight) is different than the proportion of overweight females in the population. Use the critical value approach and the 0.05 level of significance. Perform the test manually after using Minitab to summarize the data (*if you use descriptive statistics, the mean coded value in each sample is the sample proportion*).
- Explain how to find the p-value manually (indicate what probability needs to be found).
- Finally calculate manually the 95% 2-sided confidence interval for the true difference between the proportions of overweight males and overweight females.
- Explain how the results in parts b and c are consistent with your conclusion in part a.

Test and CI for Two Proportions: OWmale, OWfemale

Event = 1

Variable	X	N	Sample p
OWmale	24	39	0.615385
OWfemale	14	34	0.411765

Difference = p (OWmale) - p (OWfemale)
Estimate for difference: 0.203620
95% CI for difference: (-0.0215017, 0.428742)
Test for difference = 0 (vs not = 0): Z = 1.74 P-Value = 0.082

5 marks:

-1 mark for statement of hypotheses separately from Minitab output

-1 mark for pooling proportions $p\text{-bar} = (24+14)/(39+34)=38/73$

-1 mark for z-statistic = $(\pm 0.2036)/\sqrt{((38*35)/(73*73))*(1/39+1/34)}$

= $0.2036 / 0.1172 = \pm 1.74$ (must show manual calculation)

-1 mark for decision not to reject null H

-1 mark for conclusion that there is insufficient evidence to show proportion of overweight males is different from proportion of overweight females.

The coding in the data was incorrect because it used the 24.5 cutoff. The proper coding using the 25.5 cutoff would result in the following output:

Event = 1

Variable	X	N	Sample p
OWm_rev	19	39	0.487179
OWf_rev	12	34	0.352941

Difference = p (OWm_rev) - p (OWf_rev)
Estimate for difference: 0.134238
95% CI for difference: (-0.0902859, 0.358763)
Test for difference = 0 (vs not = 0): Z = 1.16 P-Value = 0.247

Here \bar{p} is $(19+12)/73 = 31/73 = 0.425$

And $z = (0.1343)/\sqrt{(31*42/(73*73) * (1/39+1/34))} = 0.1343/0.116 = 1.16.$

(b) p-value = $2*\text{Prob}(Z>1.74) = 2*0.0409=0.08$ or $p=2*\text{Prob}(Z>1.16) = 2 * 0.123 = 0.246$

2 marks:

-1 mark for showing one tail probability

-1 mark for multiplying this by 2.

**c. $(0.6154 - 0.4118) \pm 1.96 * \sqrt{0.6154*0.3846/39 + 0.4118*0.5882/34}$
 $= 0.2036 \pm 1.96 * 0.1149 = 0.2036 \pm 0.2251 = (-0.0215, 0.4287)$**

**or: $0.1343 \pm 1.96*\sqrt{0.487*0.513/39 + 0.353 * 0.647/34}$
 $= 0.134 \pm 1.96 * 0.115 = 0.134 \pm 0.225 = (-0.09, 0.36)$**

-1 mark for use of appropriate formula

-1 mark for correct z-value and correct SE calculation of 0.1149 or 0.115.

- (d) Since the p-value > 0.05 and the 95% CI does cover zero, we do not reject the null hypothesis and cannot conclude that the proportions of overweight males and of overweight females are different.

1 mark:

0.5 for mentioning each decision rule

3. [10 marks]

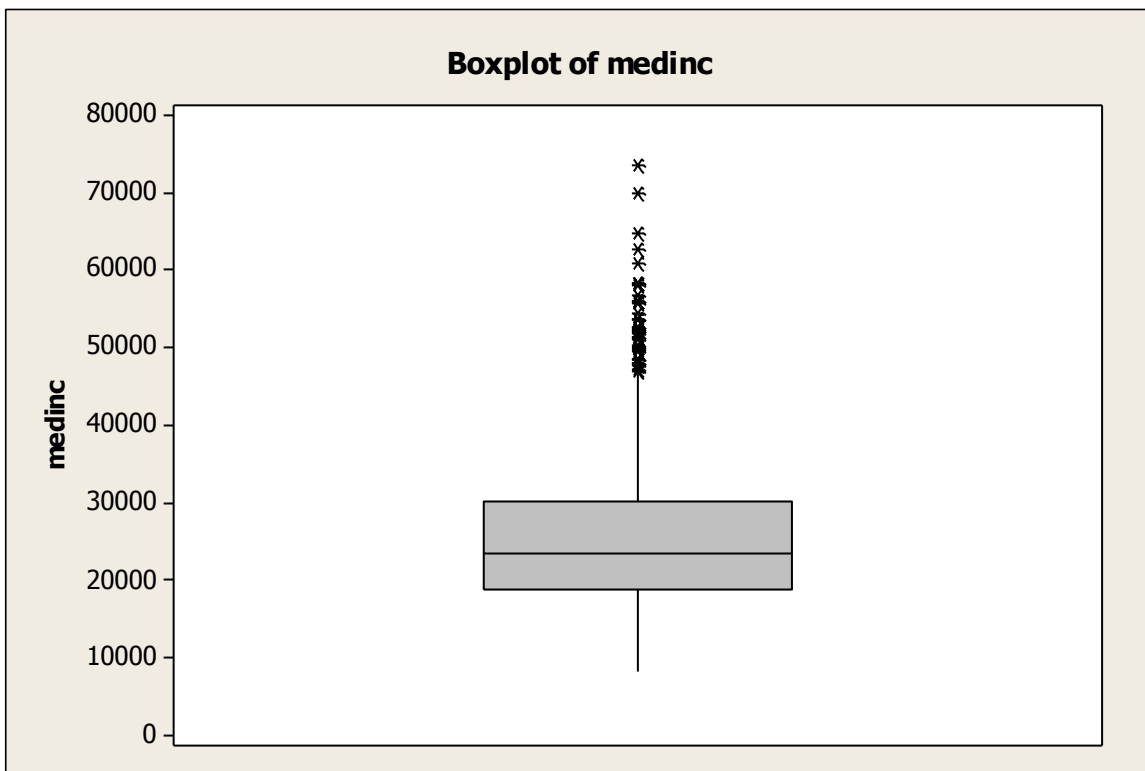
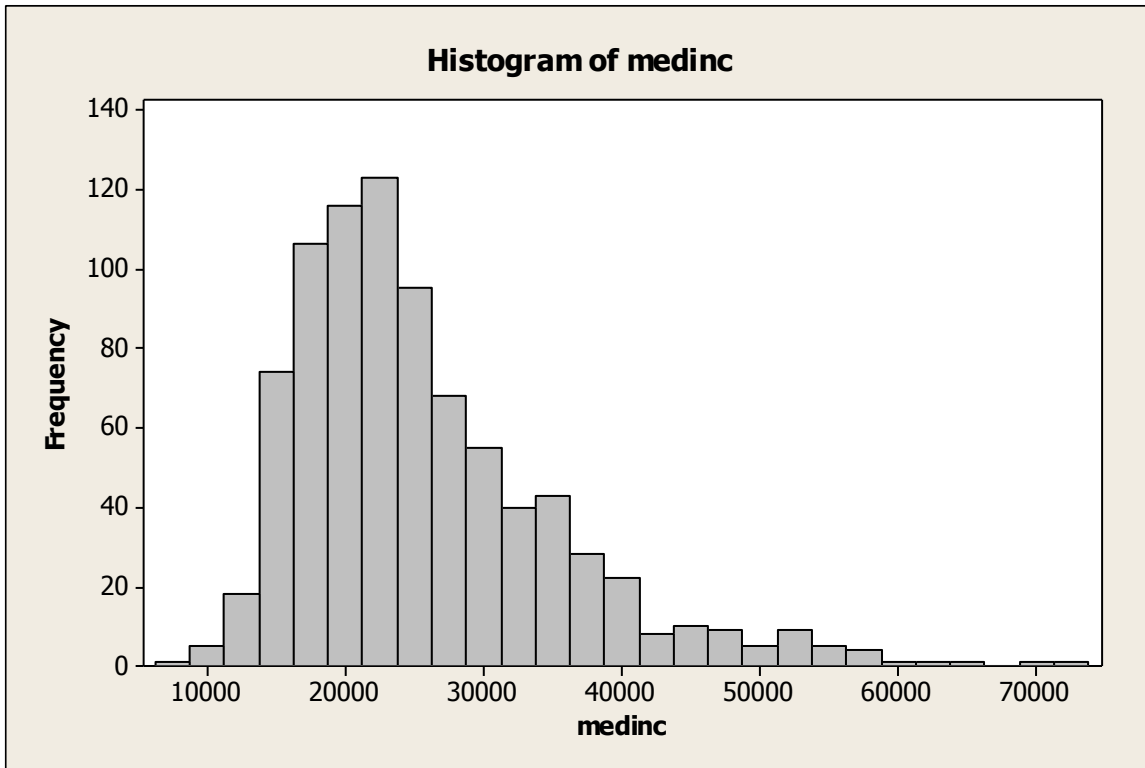
The file **Toronto.mtw** contains data on the median incomes (**medinc**) of census dissemination areas in the Toronto Census Division.

- (a) Treating this set of data as the population, use Minitab to calculate the population mean for the **medinc** variable. Set aside all population information until part e.

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
medinc	849	7	25614	327	9529	8332	18924	23429	30127

The population mean is \$25614. (1 mark)

- (b) Examine a boxplot and histogram of the population data. Explain if the mean of a random sample of size 31 would have a normal distribution.



If the judgement is that the population is not extremely skewed, then the student should say the sample mean has a normal distribution.

If the judgement is that the population is extremely skewed, then the student should say the sample mean is not normally distributed.

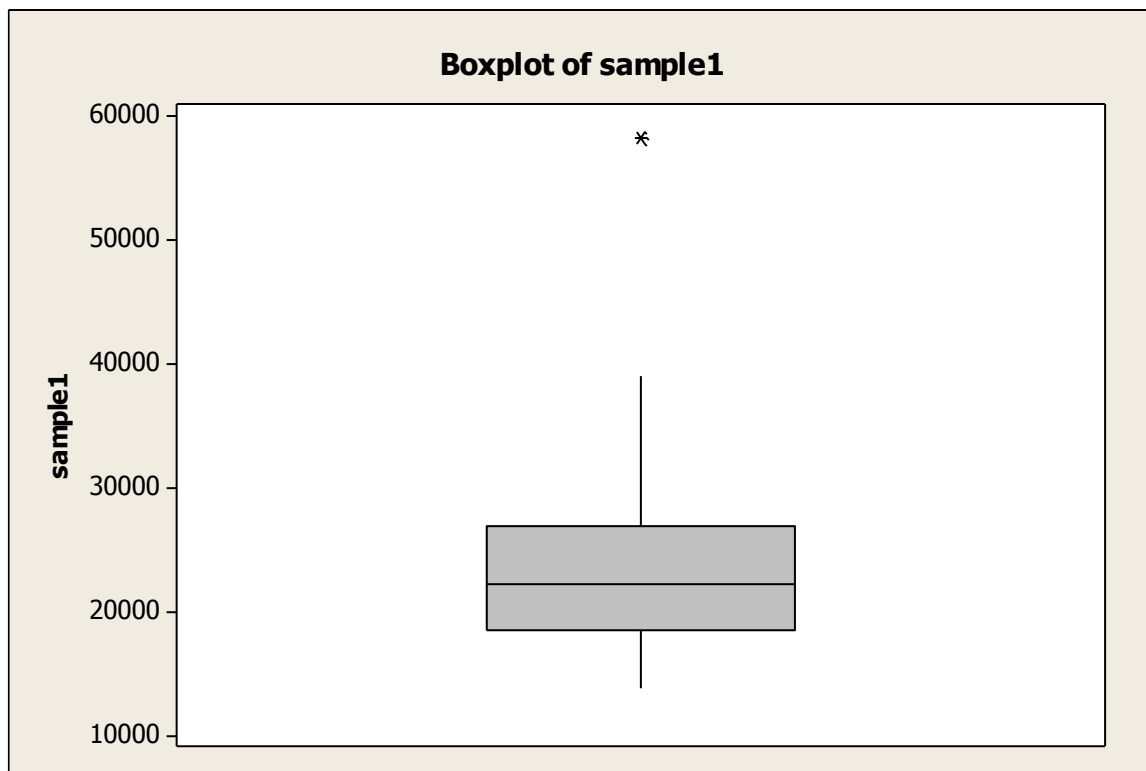
2 marks:

1 mark for at least one graph and 1 mark for one of the above comments

- (c) Now use Minitab (Calc Menu – Random Data – Sample from Columns) to draw twenty samples of size $n = 31$ from the Toronto **medinc** population. This procedure must be replicated twenty times (note that if you open up the same sampling dialog box each time from the menu, then you only have to replace the last destination column with the next one). Using each sample, use Minitab to calculate a 95% confidence interval estimate for the population mean, assuming you do not know the population standard deviation (this interval estimation can be done in one operation on all twenty columns, but not in Minitab Express).

3 marks for showing the twenty CIs

- (d) For the first sample, confirm the Minitab generated interval by calculating the interval manually. Display the sample data graphically and comment on whether the relevant assumption regarding the population distribution is warranted given your sample (state clearly the assumption needed to justify the interval estimation).



If the sample size is 31, then the relevant assumption is that the population is not extremely skewed.

If the sample size is smaller due to missing values, then the relevant assumption is that the population is approx. normally distributed.

3 marks:

-1 mark for graph

-1 marks for coming up with the correct assumption subject to the sample size

*-1 mark for **any reasonable comment** on whether this assumption is warranted (in the graph above, the assumption of a population that is normally distributed is clearly not warranted, but the assumption of a distribution that is not extremely skewed may be possible).*

- (e) Count the number of intervals out of your twenty, that contain the true value of the population mean from part a.

1 mark for the count

4. [5 marks]

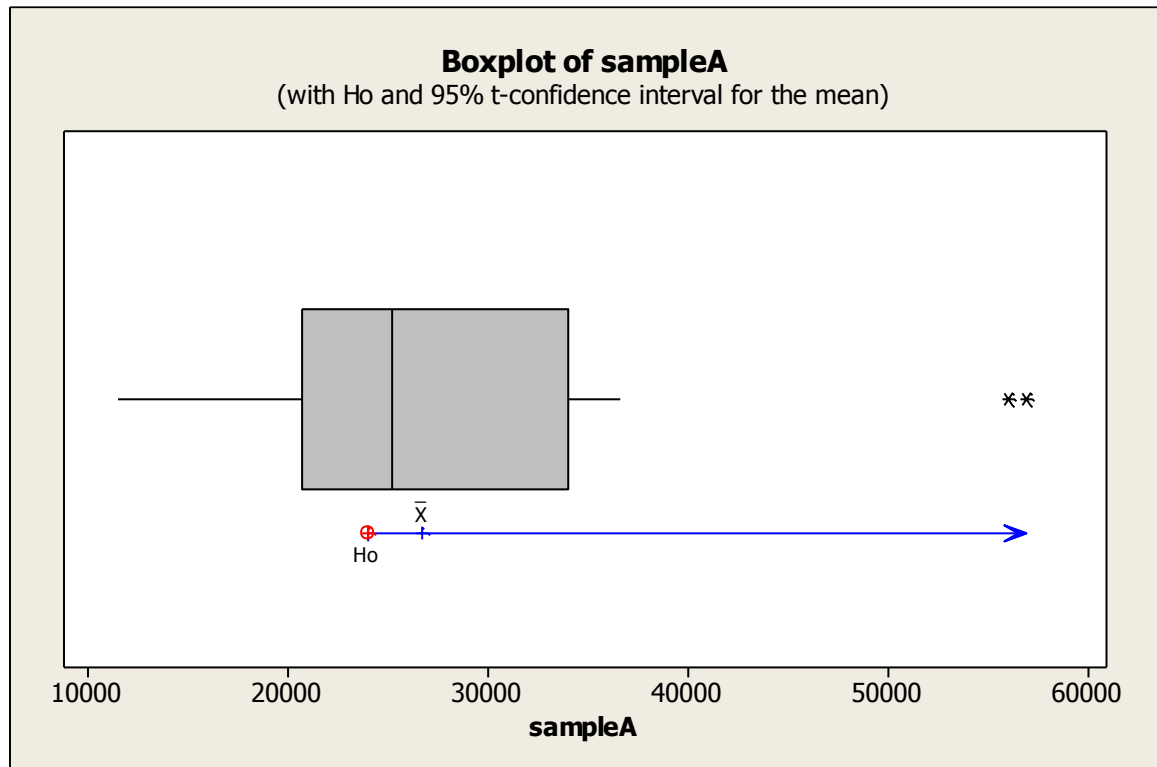
The file above also contains a random sample of size 40 (**sampleA**) from the same population of median incomes. Suppose an earlier census found that the population mean of the Toronto median incomes was \$24000. Test at the 0.05 level of significance whether the sample constitutes sufficient evidence to show that the population mean of median incomes now exceeds \$24000. Show your manual calculations (you may use Minitab to summarize the sample data).

Explain whether your test satisfies the underlying assumptions, with reference to graphical evidence.

One-Sample T: sampleA

Test of mu = 24000 vs > 24000

Variable	N	Mean	StDev	SE Mean	95% Lower Bound	T	P
sampleA	40	26705.0	10052.9	1589.5	24026.9	1.70	0.048



Since the sample size is > 30 , we only require the population not to be extremely skewed.

Although the boxplot is definitely skewed, the population may not be extremely skewed. Therefore it is safe to assume that the sample mean is normally distributed. But also accept the comment that the boxplot shows extreme skewness, therefore the sample mean may not be normally distributed.

5 marks:

-1 for hypotheses

-1 for showing $t = (26705 - 24000)/1589.5 = 1.70$

-1 for decision to reject H_0 and conclusion that the mean median income exceeds 24000

-1 for stating the assumption of a population that is not extremely skewed. Do not accept the assumption that the population is approx. normally distributed as this is not necessary given a large sample size

-1 for comment on reasonableness of assumption that the population is not extremely skewed. If the stated assumption is that the population is normally dist'd (even though this is incorrect), then accept the comment that this is not reasonable.